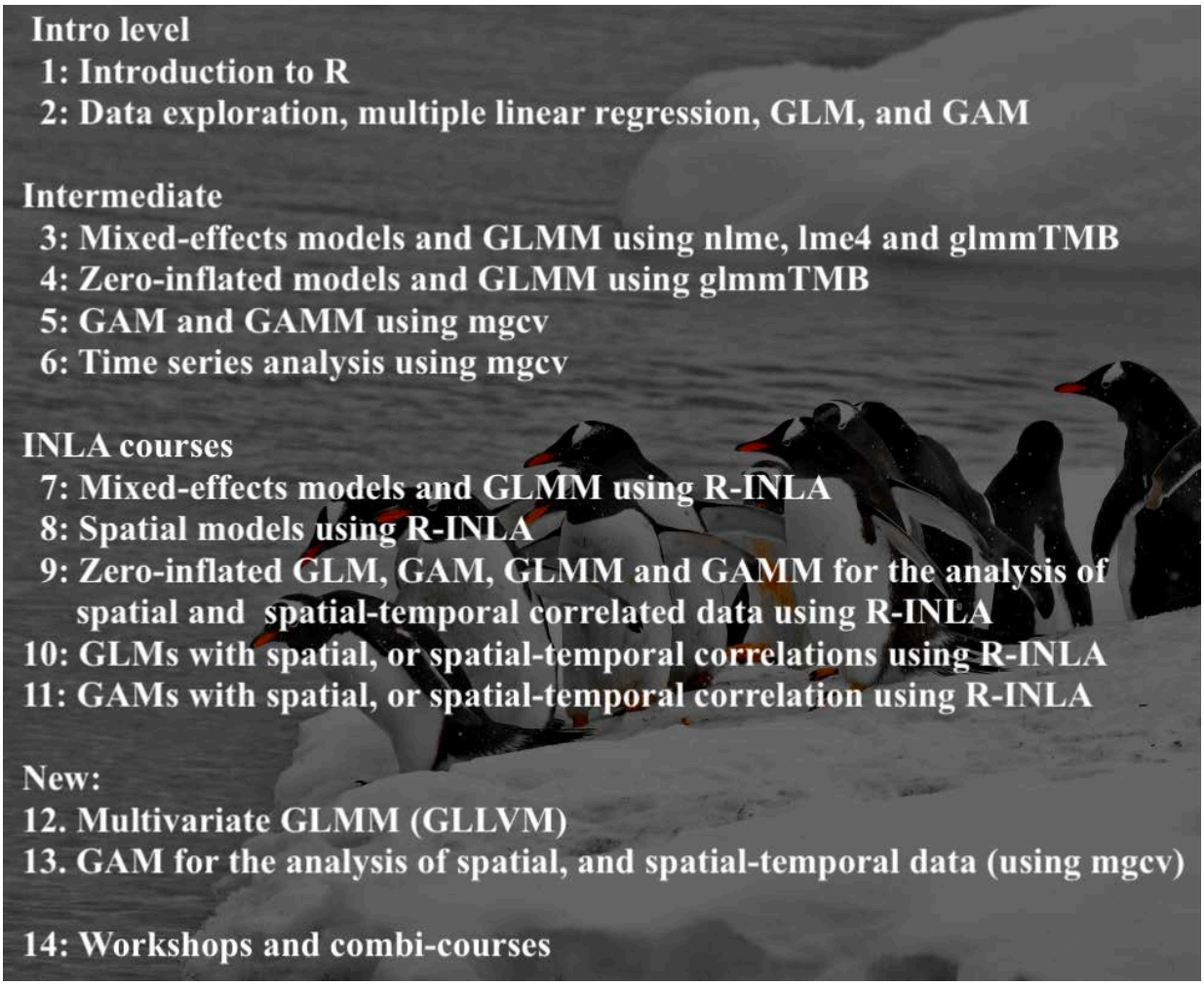
HIGHLAND STATISTICS COURSES



We provide a large number of statistics courses. The course instructors are Dr. Alain Zuur (statistician) and Dr. Elena Ieno (biologist). Our statistical and biological backgrounds ensure a lively and enjoyable interaction with our participants. One of our strong points is explaining statistics in a non-technical and understandable language.

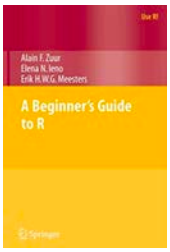
All our courses are **hybrid**, meaning they can be attended in person or simultaneously via Zoom. For a list of upcoming courses, see: http://highstat.com. Additionally, most courses are also available for self-study with on-demand videos.

If you have various colleagues who are interested in one of our courses it may be more cost-effective to organise a course at your institute. This can be done as an in-house course or as an open course. With an in-house course, you decide who participates and we charge a fixed fee. Our fee will depend on the country. For an open course, we will require a conference room (plus projector) for about 30 people, and an additional 10 local participants.

SHORT COURSE DESCRIPTION

See https://www.courses.highstat.com/ for a more detailed outline of our courses.

Course 1: Introduction to R using a protocol for conducting and presenting results of regression-type analyses



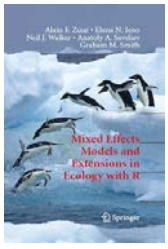
In this course, we provide an introduction to R and at the same time explain how to conduct data exploration, apply (simple) linear regression models, communicate results, and also determine optimal sample size (using power analysis) in case you want to set up a new field study or experiment.

Course 2: Data exploration, regression, GLM and GAM: with an intro to R.

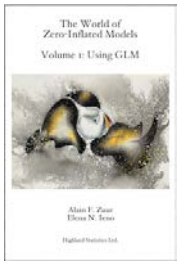
We begin with an introduction to R and provide a protocol for data exploration to avoid common statistical problems. We will discuss how to detect outliers, deal with collinearity and transformations. An important statistical tool is multiple linear regression. Various basic linear regression topics will be explained from a biological point of view. We will discuss potential problems and show how generalised linear models (GLM) can be used to analyse count data, presence-absence data and proportional data. Sometimes, parametric models (linear regression, GLM) do not quite fit the data and in such cases generalised additive models (GAM; a smoothing technique) can be used.

Course 3: Introduction to mixed effects models and GLMM

The course starts with a short revision of multiple linear regression and generalised linear models, followed by an introduction to linear mixed-effects models and generalised linear mixed-effects models (GLMM) to analyse hierarchical or clustered data, e.g. multiple observations from the same animal, site, area, nest, patient, hospital, vessel, lake, hive, transect, etc. In the second part of the course GLMMs are applied on continuous (e.g. biomass), binary (e.g. absence/presence of a disease), proportional (e.g. % coverage) and count data using the Gaussian, Poisson, negative binomial, Bernoulli, binomial, beta, and gamma distributions.



Course 4: Introduction to zero-inflated models



The course starts with a short revision of data exploration, multiple linear regression and Poisson GLM. We then discuss 3 more models for the analysis of count data, namely the negative binomial, generalised Poisson and Conway-Maxwell-Poisson GLMs. After a short theory presentation in which we explain how to extend these models towards zero-inflated models, we

apply them to various data sets. We also use the Tweedie GLM and the zero-altered Gamma GLM for the analysis of zero-inflated continuous data. In the second part of the course, we start with a short revision of linear mixed-effects models. This is followed by a series of exercises in which we analyse zero-inflated count data, continuous data, and proportional data using zero-inflated GLMMs. Throughout the course we will use the glmmTMB package in R.

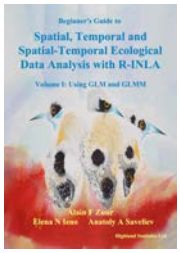
Course 5: Introduction to GAM and GAMM

We start with a short revision of data exploration and linear regression. We then introduce generalised additive models (GAM) to model non-linear relationships. We will execute these models in mgcv. In Module 2, we will revise linear mixed-effects models and show how to implement a generalised additive mixed-effects model (GAMM). We also show how to include an interaction between a smoother and a categorical covariate. In Module 3, we will revise basic GLMs and extend these towards GAMs. In Modules 4 and 5 we will discuss GAMMs for the analysis of count data, absence-presence data, proportional data, and continuous data. We also discuss 2-dimensional smoothers (including the soap-film smoother for study areas with barriers; e.g. an island in the sea).

Course 6: Time series analysis using regression techniques

The course starts with a short revision of data exploration and multiple linear regression models. A non-technical introduction of generalised additive models (GAM) is provided. GAMs will be used to estimate long-term trends, seasonal patterns, covariate effects and auto-regressive correlation. We also provide a short introduction to linear mixed-effects models and generalised linear mixed-effects models (GLMM) to analyse hierarchical data (e.g. short time series from the same core or site). GLMMs and GAMMs are used to estimated trends, seasonality and covariate effects in multivariate time series.

Course 7: Mixed-effects models and GLMM using R-INLA



The course begins with a brief revision of multiple linear regression, followed by an introduction to Bayesian analysis and how to execute regression models in R-INLA. We then explain linear mixed-effects models to analyse nested data, followed by a series of mixed modelling exercises in R-INLA. Nested data means multiple observations from the same animal, site, area, nest, patient, hospital, vessel, lake, hive, transect, etc. In the second part of the course GLMMs are applied on count data, binary data (e.g. absence/presence of a disease), proportional data (e.g. % coverage) and continuous data (e.g. biomass or distance) using the Poisson, negative binomial, Bernoulli, binomial, beta and gamma distributions.

Course 8: Intro to Regression Models with Spatial Correlation using R-INLA

We begin with an introduction how to add dependency to regression models using frequentist tools. After discussing the limitations of this approach we switch to Bayesian techniques. R-INLA is used to implement regression models, generalised linear models (GLM) and generalised linear mixed-effects models (GLMM) with spatial dependency

Course 9: Zero-inflated GLM, GAM, GLMM and GAMM for the analysis of spatial and spatial-temporal correlated data using R-INLA

We will start with a short revision of multiple linear regression, followed by a basic introduction to Bayesian analysis, and we show how to execute a linear regression model in R-INLA. In the second module, we will explain how to deal with zero-inflated count data and zero-inflated continuous data using zero-inflated Poisson, zero-inflated negative binomial, and zero-inflated Gamma GLMs. We also explain hurdle models. In the third module, we will introduce generalised additive models (GAM) to model non-linear relationships. We show how to execute these in mgcv and also in R-INLA. In the fourth part of the course, we will revise linear mixed-effects models and implement these in R-INLA. We also apply generalised linear mixed-effects models (GLMM) and generalised additive mixed-effects model (GAMM) in R-INLA. In the fifth part of the course, we will apply zero-inflated GAMs and GAMMs (and GLMMs) on various spatial correlated data sets. In module 6, we apply GAM, GAMM, and GLMM on spatial-temporal correlated data. We also deal with natural barriers for the spatial correlation (e.g. benthic species that live on a coral reef around an island). We will use barrier models; these ensure that spatial correlation seeps around a barrier (in this case an island). All exercises are executed in R-INLA.

Course 12: Multivariate GLM and GLMM using GLLVM

This course offers a journey through classical multivariate analysis techniques and advances into recently developed tools for multivariate GLM and GLMM.

We start with the multivariate techniques principal component analysis and redundancy analysis. We then continue with generalised linear latent variable models (GLLVM). A GLLVM is a GLM (or GLMM) in which multiple response variables are analysed simultaneously, while allowing for dependency between the response variables and also between the observations. We will discuss extensions of GLLVM that allow for constrained latent variables (concurrent ordination) and spatial and temporal dependency structures.

Course 13: GAM for the analysis of spatial and spatial-temporal data

We will start with a non-technical introduction to generalised additive models (GAM). Using a series of exercises, we show how GAMs can be used to allow for non-linear covariate effects. Once we are familiar with GAM, we will apply them to various spatial,

and spatial-temporal data sets. During the course, GAMs are applied to count data, absence-presence data, proportional data, and continuous data using the Gaussian, Poisson, negative binomial, Bernoulli, beta, gamma and Tweedie distributions. We will apply GAMs with 2-dimensional smoothers to analyse spatial data. To allow for natural barriers (e.g. an island in the sea), soap-film smoothers are used. On the 4th day of the course, spatial-temporal data sets are analysed.

Course 14: Workshop and combi-course

Combine the appropriate modules and use your own data sets during the course.

RECOMMENDED ORDER OF COURSES

If you do not have spatial or temporal data, then we recommend to attend the following courses within a time span of 3 years.

1. Introduction to data exploration, regression, GLM and GAM. With introduction to R (course 2).
2. Introduction to mixed effects models and GLMM (course 3).
3. Depending on your data's characteristics—such as zero-inflation or non-linear relationships—you may choose to attend either the 'Introduction to Zero-Inflated Models' or the 'Introduction to GAM and GAMM' courses (courses 4 and 5).

If you have time series data, spatial data, or spatial-temporal data and you want to use frequentist tools (e.g. mgcv), then we recommend the following two courses (in addition to course 2).

1. Time series analysis using regression techniques (course 6).
2. GAM for the analysis of spatial and spatial-temporal data (course 11).

If you are working with spatial, or spatio-temporal data and wish to use R-INLA, we recommend the following two courses (in addition to course 2). R-INLA offers greater flexibility than the mgcv approach, though it is slightly more complex.

1. Introduction to Regression Models with Spatial Correlation using R-INLA (course 8).
2. Zero-inflated GLM, GAM, GLMM and GAMM for the analysis of spatial and spatial-temporal correlated data using R-INLA'.

If you need to analyse multiple response variables at the same time, then we strongly recommend the GLLVM course in addition of courses 2 and 3.

1. Multivariate GLM and GLMM using GLLVM.

GENERAL INFORMATION

All courses are non-technical, are taught in R, and use a course website that contains:

- 5 - 10 theory presentations (with downloadable pdf files).
- 15 - 25 exercises (with downloadable data sets and documented R code).
- A Discussion Board where you can ask course-related questions.
- Live chat facilities for short questions.
- **Self-study courses**: On-demand video of all presentations and exercises, and recorded Zoom sessions from previous 'Live' online courses are available to watch.
- **Live courses**: Most live courses also contain on-demand video.

Access to the course website is for 12 months.

Live interaction.

- All courses include a 1-hour face-to-face video chat with the instructors. You can ask questions about your own data.
- Live interaction needs to take place within 12 months after being given access to the course website.

- You can start a self-study course at any time.
- The workload of each course is about 40 hours.
- A certificate (pdf file) will be provided upon completing the course. Some courses contain a short online assessment.

For further information go to: http://highstat.com

Email: Dr. Alain F. Zuur at highstat@highstat.com